

P R O J E C T M E M O R A N D U M

RAND

*Current Problems with
Army Data: Lessons from
Previous RAND Research*

Lionel A. Galway

PM-713-A

September 1997

Prepared for the U.S. Army

RAND's project memoranda are informal communications between RAND projects and their sponsors. They are intended to be timely and useful. Project memoranda have not been formally reviewed or edited. They should not be cited, quoted, reproduced, or retransmitted without RAND's permission.

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE SEP 1997		2. REPORT TYPE		3. DATES COVERED 00-00-1997 to 00-00-1997	
4. TITLE AND SUBTITLE Current Problems with Army Data: Lessons from Previous RAND Research			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Rand Corporation,1776 Main Street,Santa Monica,CA,90401-3208			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 8	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

PREFACE

The project "Logistics Information Requirements and Quality" discussed in this Project Memorandum was sponsored by the Logistics Support Activity in the U.S. Army Materiel Command and by the Director for Plans and Operations, Office of the Deputy Chief of Staff for Logistics within the Army Staff. It was conducted in the Arroyo Center's Military Logistics Program. The Arroyo Center is a federally funded research and development center sponsored by the United States Army.

OVERVIEW

Army data problems are extensive, long-standing, and hard to solve. Recent press coverage indicates that the Army's approach to solving these problems ignores some fundamental characteristics of the problems. Past RAND research suggests that much better two-way communication between the "wholesale" Army (i.e. the Army Materiel Command and Army Staff) and the "retail" Army (operational units in the field) is essential to solving data quality problems permanently.

INTRODUCTION

An article in a recent issue of *Inside the Army* reported that due to errors in entering or reporting flying hours, the Army had overstated its flying hour costs by up to \$100 million.¹ The cause of the problem is controversial: some sources say that units never reported the data, others that the Logistics Support Activity (LOGSA) did not enter the data correctly. All agree that flying hours were significantly underreported, thereby inflating the cost per flying hour as computed by the Army Cost and Economic Analysis Center (CEAC).

LOGSA can clearly correct such problems as exist with its own data entry. However, LOGSA has no direct authority to require units either to report or to resubmit erroneous reports. Therefore, according to the article, the fix being contemplated is to have LOGSA correct flying hour data from non-reporting units by using an average from similar units instead of simply putting in zero.

Such a correction procedure may be statistically questionable.² But more importantly, attempting to correct data at the point of data storage instead of at data entry eliminates the most effective and

¹Cahlink, George, "Army Finds It Has Overstated Flying Hour Costs By Up To \$100 Million," *Inside the Army*, August 18, 1997, p.2.

²See, e.g. Little, Roderick J. A., and Donald B. Rubin, *Statistical Analysis with Missing Data*, Wiley, New York, 1987. The contemplated procedure will not work if the non-reporting units differ systematically from the reporting ones in characteristics that affect flying hours. For example, deployed units (who are operating at an increased tempo) may not report flying hours as well as when they are in garrison. With the proposed procedure, the flying hours would *still* be underreported (although not as much as when using zero flying hours).

efficient method of assuring data quality. In 1996, RAND published a study of data quality problems with Army logistics data.³ This study suggests that the reason for long-term problems (and a corresponding history of failed solutions) with data from Army units that is used by wholesale organizations such as the Army Materiel Command (AMC) and the Army Staff is that attempted solutions have not focused on the *organizational* reasons why such data is not captured correctly at the point of data entry.

RESULTS OF RAND LOGISTICS DATA QUALITY STUDY

The RAND study on data quality in Army logistics (1993-1995) was prompted by numerous anecdotes about serious errors in Army logistics data systems that were having detrimental effects on the ability of Army logistics organizations (both wholesale and retail) to do their jobs. These stories were of particular concern because of the Army's continued goal of using data as a key enabler in all aspects of operations.⁴

The project looked in detail at problems with a number of key data elements, all of which had serious missing data problems. These long-standing problems had prompted a number of efforts to fix the elements, many focused at the level of the reception of the data from the units. In spite of this history of effort, the project team discovered that in one case the unit-level logistics automation software was actually discarding valid data. Although this behavior was well-known to many senior unit supply personnel and to the software developers, analysts at AMC's Major Subordinate Commands (MSCs) and LOGSA were not aware of the problem, nor had they been alerted by the developers, because collection of this data was low priority to both the units and developers.

Based on their research, the project team came to a number of conclusions. First, the Army did not fully appreciate that data was an asset, i.e. it costs money to acquire, store, and use, and this money is

³Galway, Lionel and Christopher Hanks, *Data Quality Problems in Army Logistics*, MR-721-A, RAND, 1996. This is the final report for the project "Logistics Information Requirements and Quality."

⁴Initiatives such as Force XXI and Velocity Management explicitly noted how the timely use of accurate data could substantially improve both operational and logistics effectiveness and efficiency.

well-spent only if the data are both usable (good quality) and actually used (in making decisions). Second, the problems with these data elements were not technical problems with data capture or data entry (although technical solutions had their place), but reflected the reality that much logistics data collection did not immediately benefit the units who created the data. Instead, many data elements were used only by various wholesale organizations such as AMC's MSCs for long-term studies of equipment, budgeting, etc. where the benefits accrued over several years. The units, however, directly bore the immediate costs of data collection.

This gap between where the cost of collecting data is felt and where the immediate benefits are seen greatly reduces the units' incentives to enter these elements correctly.⁵ The project's report concluded that, given that high quality data was necessary for the wholesale organizations in carrying out long-range planning and materiel improvement, data quality should be approached in terms of agreement between the chain of command of the units⁶, as data creators, and the wholesale organizations, as data users, in which the wholesale organizations justified the collection of the data by demonstrating actual use and benefit⁷ and the units accepted responsibility for assuring data quality in logistics data. Actual mechanisms for achieving this agreement were explored briefly, and included the possibility of having the data users pay for quality data.⁸

⁵The incentives are further reduced by the refusal to make some critical data elements mandatory, and by a lack of feedback to the units when data is incorrectly entered.

⁶This is somewhat ambiguous because units change commands when deployed under the control of CINCs. But it would be reasonable for FORSCOM and TRADOC to assume this responsibility in garrison.

⁷This justification would not consist of simply an *a priori* assertion of need, but also periodic review and reassessment which would include specifics of how the data is being used and what specific benefits were realized.

⁸While somewhat unorthodox, it should be noted that up until 1994 ATCOM paid contractors to collect a range of data from selected aviation units outside of the usual channels for Army maintenance data, primarily because they had more control over the data's quality.

RECOMMENDATIONS

The *Inside the Army* article demonstrates that data problems can have a major impact on the Army's operations and planning (and the credibility of the latter). The data quality project suggests that fixing problems at the level of the data archiver is not the best way to achieve data quality. While the archiver's operations do need to be high quality, and there are a number of edit checks that the archiver can do to detect quality problems, attempting to fix data at that point requires strong assumptions and often must rely on yet other questionable data.

Instead, FORSCOM/TRADOC (acting for their units and the ones which are deployed under the command of the CINCs) need to reach an agreement with the wholesale organizations which

- holds wholesale organizations responsible for justifying the collection of data not applicable to the immediate needs of the units by demonstrating the uses and benefits of the collected data. Data that are unusable or not of benefit should not be collected.
- holds units responsible for the quality of data needed to budget, plan and run operations and logistics activities (this includes giving units feedback on quality problems and insisting that data entry and correction as burden-free as possible⁹).

This would place the responsibility for balancing data burden and data benefits on the chain of command of the Army's retail units, who both bear the cost and have the ultimate benefit. The burdens and benefits of data collection need to be periodically reassessed and explicitly balanced to the satisfaction of all parties if the Army is to make real progress in solving its data quality problems.

⁹Evolving logistics automation and good logistics communication should continue to reduce the burden of collecting and archiving virtually any data needed. However, most data do require some human action to collect, and this step needs to be done conscientiously.

